TRL Limited

PUBLISHED PROJECT REPORT PPR011

OCCLUSION: PRACTICAL ASPECTS OF THE OCCLUSION METHOD TO MEASURE VISUAL LOADING AND TASK INTERRUPTABILITY

Version: 1.0

by N Brook-Carter, A Stevens, N Reed, S Thompson (TRL Limited)

Prepared for: Project Record: Occlusion UG407
Client: G Rai, VTS6, Department for Transport

Copyright TRL Limited July 2004

This report has been prepared for Department for Transport, Vehicle Telematics and Standards Division. The views expressed are those of the author(s) and not necessarily those of the Department.

Published Project Reports are written primarily for the Customer rather than for a general audience and are published with the Customer’s approval.

<table>
<thead>
<tr>
<th>Approvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Manager</td>
</tr>
<tr>
<td>N Brook-Carter</td>
</tr>
<tr>
<td>Quality Reviewed</td>
</tr>
<tr>
<td>A Stevens</td>
</tr>
</tbody>
</table>
This report has been produced by TRL Limited, under/as part of a Contract placed by DfT. Any views expressed are not necessarily those of the DfT.

TRL is committed to optimising energy efficiency, reducing waste and promoting recycling and re-use. In support of these environmental goals, this report has been printed on recycled paper, comprising 100% post-consumer waste, manufactured using a TCF (totally chlorine free) process.
Executive summary
Concerns exist surrounding the safety of operating in-vehicle information systems (IVIS) whilst driving. Although these systems can provide the driver with a vast array of useful information, they also have the potential to distract the driver by imposing high levels of visual loading. This can be a particular problem if the IVIS tasks require sustained and uninterrupted attention.

Occlusion is a technique that investigates whether in-vehicle tasks can be carried out in short bursts of visual attention towards a display screen (typically 1 to 2 seconds). This enables the tasks to be assessed in terms of both visual loading and interruptability. Safety margins or limits for the amount of visual loading a task imposes can then be investigated using the technique. Occlusion is either performed using a shutter to hide and expose the IVIS from view, or goggles that also block or reveal the IVIS. The amount of time (duration and frequency) that the IVIS is visible or occluded (blocked from view) is controlled during measurement.

A trial was carried out under the Occlusion project UG407, funded by the Department for Transport. The aim of the project is to identify gaps in the occlusion research during a literature review and to then conduct research addressing some of these key issues. This trial was designed to address the following key issues, identified at an earlier stage within the project:

- Amount and type of training
- Number of repetitions during testing
- Sample size required
- Dealing with system response delay

Training levels and the number of repetitions is investigated in order to establish the effect their variation has on results and how they should be controlled during occlusion studies. Sample size is also investigated in order to establish the appropriate sample for robust and repeatable results.

The trial involved twenty participants. Individuals considered as ‘technophobes’, or those who were highly experienced with route guidance systems and new technology, were excluded from the study. Participants’ ages ranged from 28 to 55 years.

When drivers interact with in-vehicle information systems they may experience a system delay as a result of information retrieval or computational calculations. These delays will affect the results of the occlusion tests. Three different methods of dealing with system delay are investigated in this study in order to determine how delays should be accounted for when assessing a task using occlusion.

Participants were required to carry out the following IVIS tasks:

- Task 1. Destination entry
- Task 2. Place of interest search
- Task 3. Changing radio frequency
- Task 4. Dialling a telephone number

Participants received five training sessions on these tasks (some wearing the occlusion goggles, others with no occlusion) and were then required to perform the tasks three times whilst their vision was periodically occluded by PLATO Occlusion goggles. The total task time without occlusion - Total Task Time static (TTTstatic), Total Shutter Open Time (TSOT) and errors were recorded during all of these sessions.

The main findings from the trials were:

Concerning the IVIS tasks

- The only task considered acceptable based on an existing Japanese guideline (JAMA) was dialling a familiar telephone number in a legally installed handset. However, this used a
slightly different method and a suitable limit with this occlusion protocol remains to be agreed

- Entering a destination was the task identified as requiring the most visual attention
- Destination entry and dialling a phone number are mainly visual tasks

**Concerning the Method**

- Five training sessions are required for complex tasks although fewer would be needed for less complex or familiar tasks. Task complexity is important when choosing an appropriate number of training sessions
- IVIS task training can be conducted equally effectively with or without the occlusion goggles
- Inconsistent tasks will require a greater numbers of test sessions to obtain a representative task time
- An appropriate sample size is dependent on the variability of the task. For the tasks investigated an appropriate sample size would be 14. Alternatively a smaller sample size could be used and the number of test sessions could instead be increased. Five test sessions with 10 participants appears reasonable (although not directly supported by statistical rigour)
- It is useful to remove system response delay from Total Task Times or a proportion of system response delay from Total Shutter Open Times, but the addition of a correction factor for short “check glances” is less useful
- Errors were not a useful measure of visual loading
- TSOT appears to be a good indicator of the visual loading imposed by an IVIS task
- The TSOT/TTTstatic ratio appears to be a good indicator of the amount of visual loading imposed by a task in relation to other sensory channels and provides an indication of interruptability

**Recommendations for the protocol**

- Five training sessions should be completed
- Five test sessions should be completed
- A sample size of 10 should be used
- TSOT and the ratio TSOT/TTTstatic should be used as measures of visual loading and interruptability
- Errors should not be used as a measure of performance
1 Introduction

In-vehicle information systems (IVIS) can provide the driver with a vast array of useful information. Navigation systems, road traffic information systems, mobile telephones and in-vehicle entertainment systems are all becoming widely available to the consumer. However these secondary tasks can also distract the driver from the primary task of safely controlling the vehicle. The level of distraction imposed by an IVIS task is dependent on the visual demand of the task, the interruptability of the task, and the driver’s behaviour. Although it is difficult to control driver behaviour, the visual demand and interruptability associated with an IVIS task can be controlled for through the system’s design.

Occlusion is a technique that can be used to assess both the visual demand and interruptability of an IVIS task. The technique is used for simulating the shift in visual attention associated with driving where a driver has to share visual resource between the driving scene and the in-vehicle information system (IVIS). This is either done using a shutter to hide and expose the IVIS from view, or goggles that also block or reveal the IVIS. The amount of time (duration and frequency) that the IVIS is visible or occluded (blocked from view) is controlled during occlusion. The objective is to investigate whether in-vehicle tasks can be carried out in short bursts of visual attention towards a display screen (typically 1 to 2 seconds).

This report details the methods and findings from a trial carried out using the occlusion method measuring the visual load imposed by IVIS. The trial was carried out under the Occlusion project UG407, funded by the Department for Transport. The aim of this project is to identify gaps in the occlusion research during a literature review and to then conduct research addressing some of these key issues. This trial was therefore designed to address the following key issues, identified at an earlier stage within the project:

- Amount and type of training
- Number of repetitions during testing
- Sample size required
- Dealing with system response delay

Training levels and the number of repetitions is investigated, in order to establish the effect their variation has on results and how they should be controlled during occlusion studies. If, as it is hypothesized, different levels of training and number of repetitions significantly affects the outcome of occlusion studies, then it is fundamental that a proposed occlusion protocol controls these factors appropriately.

Sample size is also investigated in order to establish the appropriate sample for robust and repeatable results. This is a particularly vital issue; if a large sample size is required for robust, repeatable results then the method will not be a cheap, low resource method for assessing IVIS. It will therefore not appeal to manufacturers.

When drivers interact with in-vehicle information systems they may experience a system delay as a result of information retrieval or computational calculations. These delays will affect the results of the occlusion tests. Three different methods of dealing with system delay are investigated in this study in order to determine how delays should be accounted for when assessing a task using occlusion

The trial described in this report involved twenty participants who carried out specific IVIS tasks. Participants received different types of training on these tasks and were then required to perform them whilst their vision was periodically occluded by PLATO Occlusion goggles.

This report details the methodology used during the trial, how the results obtained were analysed, the findings in relation to the amount and type of training required, the appropriate number of repetitions required and suitable sample sizes. The report also considers the suitability of the various variables (e.g. errors, total task times) in assessing IVIS tasks.
2 Methodology

2.1 Participants

Twenty licensed drivers from the TRL participant database were involved in this trial. Individuals considered as ‘technophobes’, or those who were highly experienced with route guidance systems and new technology, were excluded from the study. Participants’ ages ranged from 28 to 55 years (with a mean of 41.3 years, SD 9.0). Their driving and technology experience were ascertained using pre-trial questionnaires and are shown in Table 1.

<table>
<thead>
<tr>
<th>Experience (yrs)</th>
<th>Driving</th>
<th>Computer</th>
<th>Mobile telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>21.3</td>
<td>10.5</td>
<td>6.1</td>
</tr>
<tr>
<td>Minimum</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>35</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>SD</td>
<td>8.2</td>
<td>6.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Table 1 Participants’ driving and technology experience (N = 20)

Participants reported either using a stereo whilst driving frequently or very frequently. Ten of the participants reported never using a mobile telephone whilst driving, six very infrequently, one occasionally, two frequently and one very frequently. Finally, one participant reported using a navigation system whilst driving occasionally, one very infrequently and the remaining 18 participants never.

2.2 Equipment

2.2.1 Occlusion goggles

PLATO Occlusion goggles, as shown in Figure 1, were used to block participants’ vision intermittently as they completed test sessions. The goggles fit like spectacles and there is padding on the arch for the nose and on the inside of the arms to provide some comfort.

![Figure 1 PLATO Occlusion goggles](image)

The goggles are operated by a Translucent-Technologies control box linked to a computer that controls the duration and frequency of occlusion periods. The lenses are simultaneously switched from their light scattering occluding state to their transparent state in just under 1 millisecond. In the transparent state, up to 90% of incident light is transmitted. Transition back to the occluded state takes
about 5 milliseconds. When the display closes, luminance is maintained to prevent effects of light and dark adaptation.

The occlusion interval is the period of time in which the goggles are closed and vision is occluded. In relation to driving, this period is assumed to represent the time when the driver’s visual attention is directed to the surrounding road environment in the primary task of controlling the vehicle. It is also referred to as the ‘goggle close’, ‘shutter close’, or ‘display off’ interval.

The vision interval is the period of time in which it is assumed that the driver has transferred visual attention from the primary task of driving to the secondary (i.e. IVIS) task. It can also be referred to as the ‘goggle open’, ‘shutter open’, or ‘display on’ interval. Research covered in a literature review earlier in the project UG407 has investigated the interval pacing, variability and length of these occlusion intervals.

The current (pre-CD) draft of the occlusion standard (ISO N3XX) proposes a method of occlusion for assessing the visual or visual manual interfaces accessible to the driver whilst the vehicle is in motion. The draft standard states that the vision interval shall be 1.5 seconds and the occlusion interval shall be 2.0 seconds. Consequently these were the intervals used in this study.

2.3 Experimental design

2.3.1 IVIS tasks

The four tasks were:

Task 1. Destination entry
Task 2. Place of interest search
Task 3. Changing radio frequency
Task 4. Dialling a telephone number

See Appendix A for details of the information participants were asked to input/search for.

The vehicle in which participants completed the tasks was a BMW 7-Series (E38 saloon, year of manufacture 1998) and was stationary at all times. Tasks 1-3 were carried out using its factory equipped GPS-Navigation and entertainment system. Task 4 was carried out using a Nokia 7250i mobile telephone placed in a dashboard-mounted cradle, meeting current UK legislation on mobile telephone use within vehicles.

Task 1: Destination entry

Participants were given a street name and town name and asked to enter them into the navigation system as the destination.

Task 2: Place of interest search

Participants were given a specific place of interest, such as a golf course or hotel in a given town, and were asked to search for that place of interest within the navigation system and input it as the destination.

Task 3: Changing radio frequency

Participants were given a specific radio frequency and station name and asked to search for and select it using the search arrow buttons (not the pre-selected station buttons).
Task 4: Dialling a telephone number

Participants were asked to enter a telephone number that was familiar to them from memory (so that they did not have to read a number written down) into the cradled mobile telephone and select the call button.

When the trial tasks began the investigator remained quiet unless asked a specific question. An additional investigator sat in the rear of the vehicle to record task completion times and system response delays.

For each of these attempts, the observers recorded the Total Task Times, any system response delays and the number of errors made.

2.3.2 Trial groupings

All participants completed five training sessions for each of the four IVIS tasks:

- Destination entry
- Place of interest search
- Changing radio frequency
- Dialling a telephone number

For the training sessions, the twenty participants were each allocated to one of two groups. As shown in Table 2, the ten participants in Group A performed each IVIS task in the five training sessions without any form of occlusion (static training). The ten participants in Group B performed five training sessions of each of the four tasks whilst wearing occlusion goggles (occluded training). There then followed three test sessions of each task in which all participants wore occlusion goggles.

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>No occlusion</td>
<td>No occlusion</td>
<td>No occlusion</td>
<td>No occlusion</td>
<td>No occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
</tr>
<tr>
<td>B</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
<td>Occlusion</td>
</tr>
</tbody>
</table>

Table 2 Summary of trial groupings
Participants completed the IVIS tasks in a pseudo-random order to counterbalance and minimise any possible system learning and order effects. See Table 3 for the task orders for each participant.

<table>
<thead>
<tr>
<th>Participant number</th>
<th>Task 1 Destination entry</th>
<th>Task 2 Place of interest search</th>
<th>Task 3 Changing radio frequency</th>
<th>Task 4 Dialling a telephone number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3 Participant Experimental Condition task order coding

2.3.3 Variables

The independent variables were:

- the IVIS tasks
- training technique (Group A (static) vs. Group B (occluded))

The dependent variables were:

- Total Task Time static (TTTstatic)
- Total Shutter Open Time (TSOT)
- Errors
2.4 Procedure

2.4.1 Preparation
Before beginning the trial, participants were asked to read an information sheet containing instructions relating to the nature of the trial. If participants were happy with what was required of them and wished to continue with the trial, they were asked to fill in a consent form.

Participants were then administered the driver profile questionnaire in order to obtain details relating to the driver’s demographics and experience both in terms of driving and in terms of using new technology. This included the questions on how long they had used equipment such as mobile telephones and how often they would use them whilst driving.

2.4.2 Participant accommodation
Each participant occupied the driver’s seat of the vehicle throughout their test session. When the participants were carrying out the sessions wearing occlusion goggles, the goggles were connected to a laptop which was placed to the left side of the driver’s seat.

2.4.3 Training

2.4.3.1 Using occlusion goggles
All participants were provided with training in wearing the occlusion goggles before beginning the sessions in which they were required to wear them. Participants were given time to familiarise themselves with the goggle operation and asked to carry out a number of simple requests such as ‘turn on the vehicle headlights’ and ‘adjust the climate control’ while the goggles were running.

2.4.3.2 Carrying out IVIS tasks
All participants were given step-by-step verbal instructions on how to carry out each IVIS task. See Appendix A for full step-by-step instructions for each IVIS task.

Group A participants then completed five timed training sessions of each task without any occlusion. Group B participants completed five timed training sessions of each task but wore the occlusion goggles, periodically blocking their vision of the IVIS apparatus. Each participant completed sessions in the order shown in Table 3.

2.4.4 Test sessions
After completing the training sessions, all participants completed 3 timed test sessions of each task in the order shown in Table 3, whilst wearing the occlusion goggles.
3 Analysis

Total Task Times (static) were recorded during the five training sessions for the ten participants in Group A (static training). Total Task Time (static) was the time taken to complete a task when not wearing the occlusion goggles.

Total Task Times (occluded) were recorded during the study for the training sessions completed by the ten participants in Group B (occluded training) and for all participants during the three test trial times. Total Shutter Open Times were then calculated based on the Total Task Times and the vision interval.

During the trial, system response delays were timed on each occasion a participant attempted a task. A system response delay is classed as a delay in the IVIS response to driver input. This delay cannot be controlled by the driver. Often during periods of system response delay, feedback relating to the system status is provided to the driver, for example in the form of an egg timer. If no system response delay feedback is provided to the driver then it is considered appropriate to not take the system response delay into account, as the user will have to keep checking the system for a change in status and may also be unsure of the system status.

System response delays were found to vary considerably between different attempts at both the same and different tasks. This system response delay will impact on the total time taken to complete a task and consideration needs to be given as to how it can be accounted for in calculations.

Three methods were identified and are described below:

Static Total Task Times
- A) No correction
- B) Remove delay
- C) Remove delay and add time for short glances

Total Shutter Open Times
- A) No correction
- B) Remove proportion of delay
- C) Remove proportion of delay and add time for short glances

The first method (A) is the same for both Static Total Task Times and Total Shutter Open Times. This method of making no correction is based on the fact that the system did not provide participants with feedback that a system response delay was taking place in the form of either a visual cue, such as an egg timer, or an auditory cue, such as a beep to indicate the system was ready. A problem with this method is that the variation in system response delays for the same task is not removed.

The second method (B) of removing the delay altogether has the advantage of removing the “noise” caused by system response delay. It is simple to remove the delay from Static Total Tasks Times. For Total Shutter Open Times a formula has been derived for calculating an estimate of how much delay occurred during vision intervals of the occluded conditions. This proportion of delay is calculated by:

\[
\text{System response delay} \times \frac{\text{vision interval}}{\text{vision interval} + \text{occlusion interval}}
\]

One limitation of this second method is that it does not account for any participant interaction with the system during these periods of delay to check for when the system is ready. So, in the third method (C), the system response delay is removed from either the Static Total Task Time or Total Shutter open Time (as in the second method) and then time is added to account for the short glances which participants often make whilst checking for the end of the system response delay. This was estimated to involve half second glances at two second intervals, so the additional factor is:

\[
\text{System response delay} \times 0.25
\]
4 Results

Where labelled, the four tasks are:

- Task 1. Destination entry
- Task 2. Place of Interest (PoI) search
- Task 3. Changing radio frequency
- Task 4. Dialling a telephone number

4.1 Correcting for system response delay

As mentioned earlier, in correcting task times to compensate for the effect of system response delay, three methods were identified and are described in section 3 and below.

Static Total Task Times (Group A training):

A) No correction
B) Remove delay
C) Remove delay and add time for short glances

Total Shutter Open Times (Group B training and all participants test sessions):

A) No correction
B) Remove proportion of delay
C) Remove proportion of delay and add time for short glances

Figure 2 shows the effect of removing system response delay from task completion times (both Static Total Task Times and Total Shutter Open Times) using data from participants in both groups across training sessions.

![Figure 2](image)

Figure 2 Comparison of system response delay corrections for all participants on the ‘Place of Interest search’ task across training sessions.
As can be seen in Figure 2, the system response delay has had a marked effect on the training time, particularly in training sessions 3 and 4. Furthermore, the 95% confidence intervals are much tighter for the Task times with system response delay removed. This is confirmed by the reduction in the standard deviation of mean task times when system response delay has been removed. Using training session 4 as an example, the mean task time was 53.0 (N = 20, SD = 23.1). Removing the system response delay reduces the mean task completion time to 20.5s (N = 20, SD = 9.64). By contrast, the correction for short glances had little effect on the task times beyond the addition of a constant value.

For these reasons, subsequent analyses of data from Tasks 1 and 2 use task times that have had the system response delay (or a proportion of the system response delay) removed.

It can be concluded that it is appropriate to remove system response delay from TTTstatic and a proportion of system response delay from TSOT, but that adding a correction for short glances has little effect.

4.2 Training progress

Figures 3, 4, 5 and 6 show how participants performed in each of the four tasks as they completed the five training sessions. They show Total Task Time for participants in Group A and TSOT for participants in Group B. Where appropriate, the system response delay has been subtracted from the Task time (i.e. Tasks 1 and 2).

![Task 1: Destination entry](image)

Figure 3 Mean [Total Task Times – System response delay] (Group A) or Mean [TSOT – System response delay] (Group B) for participants in each Group completing the Training sessions of Task 1.
**Task 2: PoI search**

![Graph showing Mean [Task time - System delay] for different groups over training sessions.]

**Figure 4** Mean [Total Task Times - System response delay] (Group A) or Mean [TSOT - System response delay] (Group B) for participants in each Group completing the Training sessions of Task 2.

**Task 3: Change radio freq.**

![Graph showing Mean Total Task Times (Group A) or Mean TSOT (Group B) for different groups over training sessions.]

**Figure 5** Mean Total Task Times (Group A) or Mean TSOT (Group B) for participants in each Group completing the Training sessions of Task 3.
A t-test comparison revealed that the time taken by participants in group B to complete tasks 2, 3, and 4 was significantly less than that of participants in group A (df\(^1\) = 98 in each case, \(p_2 < 0.05\) (task 2), \(p < 0.005\) (tasks 3 and 4)). This is unsurprising because participants in group B can perform task non-visual processing during occluded periods and this time is not included in the TSOT. The fact there was no difference between groups A and B in task 1 suggests that it requires virtually exclusive visual attention.

The effect of the training sessions on task performance was assessed by measuring the Pearson linear correlation between Task time and training session across the two groups in the four different tasks. A significant negative correlation between Task time and Session indicates that over the sessions included in the analysis there was a significant decrease in Task time. By the successive removal of the earlier training sessions from the analysis and repetition of the correlation tests, it is possible to see whether the Task time has reached a constant level in the later trials as the significant negative correlation will disappear.

For participants in Group A, only the Training sessions can be considered (sessions 1-5) because the Training sessions were carried out under different conditions to the Test sessions (with the participants wearing the occlusion goggles). Table 4(a) shows how the correlation between Task time and Session changes as the early sessions are removed from the analysis.

---

\(^1\) df = degrees of freedom - The degrees of freedom is the number of values in the calculation of a statistic that are free to vary. In other words the total number of observations in the sample minus the number of samples.

\(^2\) p = the p value - The probability that the difference(s) observed between the two groups would have occurred if there were no differences between the groups other than those created by random selection. The assumption underlying the p-value is the null hypothesis.
Table 4(a) The loss of Pearson correlation between training session and task time across tasks for participants in Group A.

For participants in Group A, in tasks 1-3 the correlation is non-significant once the first training session is removed from the analysis suggesting that two training sessions are required for the participant to reach an optimum performance level. In task 4, dialling a telephone, there is no significant negative correlation. This is unsurprising since the task is very familiar to most people and practice is unlikely to yield any improvement in performance.

When repeating the test for participants in group B it is possible to use training and test sessions (sessions 1 to 8) since Task conditions did not vary between the Training and Test scenarios – group B participants wore occlusion goggles in all trials. Table 4(b) shows how the correlation between Session and Task time changed as training progressed.

Table 4(b) The loss of Pearson correlation between training session and task time across tasks for participants in group B.

For participants in Group B, the correlation between session and Task time remains significant for Task 2 throughout training (p < 0.005 for each correlation) until session 5. This suggests that participants learning to perform the Place of Interest search whilst wearing the occlusion goggles need at least five training sessions before they can be considered proficient at the task. For the other tasks, results suggest that participants require no more than two training sessions before reaching an optimum performance level.

It can be concluded that five training sessions would be required for complex tasks and fewer would be needed for less complex or familiar tasks. Task complexity is therefore important when choosing an appropriate number of training sessions.
4.3 Static training times

Figure 7 Graph showing Mean Total Task Times for Group A participants in (static) training sessions by system response delay correction and IVIS task (dashed line indicates SAE 2364 (2000) time limit recommendation).

Figure 7 summarises the Mean Total Task Times when vision was not occluded. Tasks 3 and 4 did not involve a system delay and only Total time is displayed. SAE 2364 (2000) recommends that the total task time for navigation system functions accessible while driving does not exceed 15 seconds. The mean total task times for tasks 1 (Destination entry), 2 (Place of interest search) and 3 (Change radio frequency) were well above the recommended limit of 15 seconds.

The different methods of correction for system response delays only affected tasks 1 and 2, as no delays were observed in tasks 3 (Change radio frequency) or 4 (Dial telephone number). Removing the system response delay entirely reduced the static TTTs for tasks 1 and 2 but they still remained above the 15 second limit. Adding a correction for glances to the IVIS during a system response delay increased the static TTTs, taking them further above the 15 second limit.

The only task that did not exceed the recommended Static Total Task Time recommended limit of 15 seconds was dialling a familiar phone number in a legally installed handset.
4.4 Occluded training times

![Graph showing Mean Total Shutter Open Times (TSOTs) for Group B participants in (occluded) training sessions by system response delay correction and IVIS task (dashed line indicates JAMA time limit guideline).](image)

The JAMA guidelines (2000) recommend that the total shutter opening time from the start to the end of the control, with vision intervals of 1.5 seconds and occluded intervals of 1.0 second, should be 7.5 seconds. The mean Total Shutter Open Times (TSOTs) for each IVIS task are shown in Figure 8, illustrating that task 1 had by far the highest TSOT, well over 7.5 seconds. Tasks 2 and 3 also both exceeded the recommended 7.5 seconds. Finally, task 4 was within the recommended limit for TSOT. It should be noted that during this study occluded intervals of 2.0 seconds were employed, giving more ‘thinking time’ between vision intervals than the 1.0 second recommended by the JAMA guidelines.

Again the correction for system response delays only affected tasks 1 and 2. Removing the system response delay entirely reduced the TSOTs for these tasks but they still remained above the 7.5 second limit. Adding on a correction for glances to the IVIS during a delay increased the static TTTs, taking them further above the 7.5 second limit.

The only task considered acceptable based on existing guidelines (JAMA) was dialling a familiar telephone number in a legally installed handset. However it should be noted that a suitable limit with this protocol remains to be agreed.
4.5 Test times

After the 5 training sessions on each task, both groups of participants completed three Test sessions of each Task whilst wearing the occlusion goggles. Consequently, TSOT times can be quoted for both participant groups. As before, system response delays have been subtracted from the times of participants in Tasks 1 and 2. Unlike the training sessions, there was no randomisation of trial order for each task i.e. each subject had to perform the same task in each of three test instances of that task – test session 1 was the same across all subjects, test session 2 was the same across all subjects, and test session 3 was the same across all drivers. Figures 9 – 12 show how participants performed through the three test sessions.

![Task 1: Destination entry](image1)

**Figure 9** Mean [TSOT – System response delay] for participants in each Group completing the Test sessions of Task 1.

![Task 2: PoI search](image2)

**Figure 10** Mean [TSOT – System response delay] for participants in each Group completing the Test sessions of Task 2.
Participants’ test performance demonstrates that, it is only in the ‘Dial a telephone number’ task that TSOTs are less than the JAMA guidelines. In all other tasks, TSOTs are considerably greater.

A t-test was performed to compare the overall Test times in each Task across Groups. The t-tests revealed that there were no significant differences between the Groups in their mean Task times. Unlike the training sessions, under test conditions Group A and Group B participants both completed the tasks wearing the occlusion goggles. The failure to detect any differences in task completion time under test conditions suggests that the training method does not have a significant effect on Test performance.

These results suggest that the type of training on the IVIS tasks (with or without occlusion) does not affect the test results.
Figures 11 and 12 suggest test instance 2 was somehow more difficult than the other test instances. However, statistical tests show that the only significant difference lies between test instances 2 and 3 in task 3. Since test instances were not randomised it is likely that the second test instance of the change radio frequency task was more slightly difficult than the third.

In Figures 9 – 12 all participants were given the same three tasks in the three test sessions. For example, the same three destinations were given for the destination entry task test sessions. This may help explain why some differences are seen between the three test sessions. For example, for the changing the radio frequency task (Figure 11) participants took longer to find the test 2 radio station (95.2 FM radio Bristol) than the Test 1 frequency (909 MW) or the Test 3 frequency (1215 MW). This might indicate that FM stations take longer to search for.

4.6 Test sessions required

To investigate the number of Test sessions that are required to obtain a representative sample of test times, a one-way ANOVA (analysis of variance) with a dependent variable of task completion time (minus System response delay for tasks 1 and 2) and a factor of Test session was performed to detect whether the mean task completion times were homogeneous across Test session. Since overall Test times showed no difference across Group, this factor was not included in the analysis. Post-hoc Tukey HSD\(^3\) multiple pairwise comparison tests were used to determine which Test sessions (if any) were significantly different from any other.

The ANOVA revealed that there was a significant variation in mean task completion time across Test sessions in Task 2 (df = 59, F = 3.69, p < 0.05) and the Tukey test showed that this difference was between Test sessions 1 and 3. There was also significant variation in mean task completion time across Test sessions in Task 3 (df = 59, F = 4.28, p < 0.05). This time the Tukey test showed that the difference was between Test sessions 2 and 3.

The statistical tests show that there were no significant differences across Test sessions for either Task 1 or Task 4 (Task 1: destination entry; Task 2: dialling a telephone number). The failure to find a significant difference across Test sessions indicates that there is a lack of variation in the amount of time taken to complete each task instance. Since participants are tending to complete these respective tasks in the same amount of time each time they perform them, there is little to be gained by making participants complete them several times. As a result, only two Test sessions are required to obtain a representative mean task completion time for these IVIS tasks.

The lack of variation in task completion time for Task 1 and Task 4 is likely to be due to the fact that there is little variability in the task demands. Entering a destination requires the participant to enter a consistent amount of text and instructing the IVIS to select a route to it. The time taken for a participant to complete this task is unlikely to vary greatly across destinations of interest. Similarly, entering a telephone number on a mobile telephone is a very consistent task requiring the participant to enter eleven digits into a familiar keypad and instruct the telephone to dial the number with a further single key-press. Consequently, it is unsurprising that no variation was observed across test instances of each of these tasks.

Conversely, in Tasks 2 and 3 (Task 2: Place of interest search; Task 3: Changing radio frequency), significant differences in task completion time were observed across Test instances of each task. This suggests that there was greater variability in the task demands than was observed for tasks 1 and 4.

\(^3\) Post Hoc Tukey test: Having determined that differences exist among the means using an ANOVA, post hoc range tests and pairwise multiple comparisons can determine exactly which of the means differ. In this case a Tukey test has been used. Comparisons are made on unadjusted values. Tukey's honestly significant difference test uses the Studentized range statistic to make all pairwise comparisons between groups and sets the experimentwise error rate to the error rate for the collection for all pairwise comparisons.
variability in changing the radio frequency is likely to be due to the participant having to change the frequency through varying amounts depending on the start and target frequencies. The variation in task completion time across the three Test instances for these two tasks suggests that it would be unwise to use fewer than three Test sessions to obtain a representative mean task completion time.

These results suggest that inconsistent tasks will require a greater number of test sessions to obtain a representative task time and that it would be unwise to use fewer than three Test sessions for tasks with high variation such as searching for a Place of Interest or a radio frequency.

4.7 Participants required

Twenty participants were involved in this study. However, the purpose of the study was to determine whether fewer than twenty participants might provide similar representative results. Therefore in order to determine the minimum number of participants that would be required to obtain results representative of the outcome achieved in these trials the following analysis was carried out.

The mean task completion times for all twenty participants’ test trials in each task were compared with the N participants who completed each task in the fastest time and also with the N participants who completed each task in the slowest time (where N is a number less than the total number of participants used in trials; in this case, twenty).

The value of N was varied systematically to find the number at which the difference in test times between the fastest N participants and all twenty participants is non-significant and the difference between test times of the slowest N participants and all twenty participants was non-significant (p > 0.05). At this point, N gives an indication of the smallest number of participants that could be used to match the results of the larger group.

As in previous analyses, a proportion of the system delay was subtracted from the TSOT times for tasks 1 and 2 (no such correction is made in tasks 3 and 4 as no system delay was encountered).

Independent-sample t-tests were used to compare the mean test times in each task across the fastest, slowest, and all trial participant groupings for the range of N values tested. The comparisons revealed that for task 1 and task 2, ten participants were required to find no significant difference between the mean test times of all participants and either the fastest ten or the slowest ten participants. For task 3, fourteen subjects were required to eradicate significant differences between the all participants group and the fastest/slowest groups. In task 4, an N value of seven was required to achieve the same result.

Table 5 shows the mean test times for the minimum N participant grouping that did not differ from the All participants group in each task.

<table>
<thead>
<tr>
<th>Task</th>
<th>Group</th>
<th>N</th>
<th>Mean test time (s)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All participants</td>
<td>20</td>
<td>47.9</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>Fastest 10</td>
<td>10</td>
<td>36.5</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>Slowest 10</td>
<td>10</td>
<td>59.4</td>
<td>18.5</td>
</tr>
<tr>
<td>2</td>
<td>All participants</td>
<td>20</td>
<td>16.5</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>Fastest 10</td>
<td>10</td>
<td>12.4</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Slowest 10</td>
<td>10</td>
<td>20.7</td>
<td>4.5</td>
</tr>
<tr>
<td>3</td>
<td>All participants</td>
<td>20</td>
<td>17.1</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>Fastest 14</td>
<td>14</td>
<td>14.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Slowest 14</td>
<td>14</td>
<td>18.7</td>
<td>4.6</td>
</tr>
<tr>
<td>4</td>
<td>All participants</td>
<td>20</td>
<td>6.3</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>Fastest 7</td>
<td>7</td>
<td>3.4</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>Slowest 7</td>
<td>7</td>
<td>10.6</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Table 5 Mean task completion times for minimum participant number matched groupings of fastest, slowest, and all participants in each task.
The results of this analysis suggest that a minimum sample size of ten participants, each completing three task replications, would be required to obtain a reasonable representation of the performance of the population for the tasks involving programming of the navigation system (tasks 1 and 2). For task 3 (changing the radio frequency), a minimum of fourteen participants completing three task replications would be required. However, for task 4 (dialling a telephone), only seven participants completing three task replications would be required.

The variation in minimum participant numbers suggests differing task demands. Dialling a memorable telephone number on a mobile telephone requires the entry of eleven digits on a familiar keypad. Task completion times are unlikely to vary greatly between participants. Hence, fewer are required to obtain a representative sample. The protocol of the trials in task 3 dictated that the requirement to change the radio frequency varied significantly between trials as participants had to cycle through different frequency ranges to reach the target station in different trials. It is likely that the greater task variability is responsible for the increase in the number of participants required to obtain a suitable sample and that the number of participants required could be reduced by standardising the task across test instances. This must be balanced against the risk of trial predictability having a detrimental effect on the validity of the results.

Based on findings from this trial, the minimum number of participants required to carry out an occlusion study would be fourteen. This is the highest number of participants necessary to perform all of the tasks undertaken in this experiment to achieve results representative of those obtained by the larger sample of twenty.

4.8 TSOT/TTTstatic ratio

In order to estimate the portion of static total task time (TTTstatic) that is combined with visual attention to the task, the ratio TSOT/TTTstatic is calculated. This ratio has been suggested to be a rough measure of the degree to which a task can be done without visual control. The smaller the deviation from 1 the more visual attention is necessary for performing the task. Ideally the ratio should be much lower than 1.0.

Figure 13 shows the mean TSOT/TTTstatic ratio for each task, using the mean TSOT from the group A test sessions and the mean TTTstatic from the last three group A training session (when a learning effect was no longer observed).

![Figure 13 Bar chart showing the TSOT/TTTstatic ratio across tasks](image)

As can be seen from Figure 13, Tasks 1, 2 and 4 had ratios far closer to one than task 3. The theory would take this as implying that Task 1, 2 and 4, destination entry, searching for a Place of Interest and dialling a mobile phone number, require more visual loading than Task 3, changing a radio frequency. This may be explained by the fact that entering destination, searching for a Place of
Interest and entering phone numbers would require the driver to look at the input controls, where as searching for a radio frequency can be performed through the auditory, as well as the visual, channel.

Based on these results the TSOT/TTT ratio appears to be a good indicator of the amount of visual loading imposed by a task in relation to other sensory channels.

4.9 Errors

Figure 14 shows the total number of Errors made by participants in each Group and in each Task.

![Bar Chart showing the Sum of the Errors made by all Participants in all Sessions in each Group across Tasks.](image)

There was no significant difference in the number of errors across Group, suggesting training wearing the occlusion goggles does not cause the participant to make more errors. However, there are clearly significant differences across Task (1 or 3 vs. 2 or 4; all df = 318, p < 0.005; all other comparisons NS). The nature of tasks 1 and 3 meant that errors were more likely as participants made incorrect key presses when entering the destination and when using the seek buttons to cycle through the radio frequencies. This accounts for the greater number of errors in these tasks.

The overall number of errors made was generally very low; approximately one error was made in every ten sessions. A significant negative correlation between number of errors and training session would indicate that participants were making fewer errors as they became more experienced in using the IVIS. No such correlation existed for any group or task and therefore it appears that the errors were genuinely random mistakes.

Errors did not provide a useful index for assessing the IVIS tasks. The pre-ISO protocol states that training should be conducted by participants until error free performance is reached. However, these results suggest that this would not be appropriate for many IVIS tasks in which few errors are ever made.
5 Discussion

The IVIS tasks

The only task that could be considered acceptable under SAE 2364(2000) and JAMA guidelines was dialling a telephone number on a legally installed handset. However, it should be noted here that different vision intervals were used in this study to those specified by JAMA. Further, a TSOT limit and the TSOT/TTTstatic limit, for a standard occlusion protocol, remain to be agreed. What was encouraging from the results of this study was that the TSOT results reflected what would have been expected for each task, based on expert opinion; for example, the more complex destination entry task was found to have a higher mean TSOT than the simpler dialling a phone number task.

Similarly the TTT/TSOT ratios reflect what would be expected. Both Tasks 1 and 4 had high TSOT/TTTstatic ratios, where as the TSOT ratios for tasks 2 and 3 were moderate. The theory would take this as implying that Task 1 and 4, destination entry and dialling a mobile phone number, require more visual loading than Task 2 and 3, selecting a point of interest and changing a radio frequency.

This can be explained by the fact that entering destination and phone numbers requires the driver to look at the input controls, where as searching for a radio frequency can be done through the auditory as well as the visual channel. Further, searching for a place of interest might require less visual loading since searching a list is less visually demanding than inputting letters or numbers. The TSOT/TTTstatic ratio results appear to reflect the visual loading imposed by a task and these findings therefore support the ratio as a useful indicator of the amount of visual loading imposed by a task in relation to other sensory channels.

The method

The results suggested that it is useful to remove system response delay (or a proportion of system response delay in occluded vision sessions) for comparisons of task times. This gives times that are more representative of participants’ performance in completing sessions and significantly reduces between-trial variation allowing better comparisons to be made. The addition of a correcting factor to allow for short glances to the IVIS is less useful.

When comparing the TSOTs collected during the training sessions from participants in Group B with the static TTTs collected during the training sessions for those in group A, the TSOTs were shorter. This can be attributed to the fact that Group B participants could perform non-visual processing during occluded periods. No difference was observed in task 1 (entering a destination). It is likely that this is because entering a destination required the most visual attention. This is supported by the fact that the destination entry task (1) had the highest TSOT/TTTstatic ratio, which is believed to indicate the relative level of visual loading.

The findings from this trial suggest that three or more training sessions would be required for complex unfamiliar tasks (PoI search in particular) for participants to reach a constant performance level. Fewer training sessions would be needed for less complex or familiar tasks, such as dialling a telephone number. The task complexity should therefore be considered when choosing an appropriate number of training sessions, to ensure that an optimum level of performance is reached.

Test performance was equivalent across the two groups which received different types of training. This suggests that the training method did not have a significant effect on test performance, and that participants can be trained in the IVIS task either with or without occlusion.

Pairwise comparisons of participants’ mean task completion times for the test sessions of tasks 2 and 3 (Place of interest search and Changing radio frequency) showed that none of the results from the three test sessions were entirely heterogeneous. Mean task completion times across test sessions were (independently) homogeneous for tasks 1 and task 4. This suggests that a minimum of two task repetitions would be required to obtain a representative sample of test performance for tasks 2 and 3, and possibly fewer for tasks 1 and 4. The result is perhaps indicative of the consistency of task demands. The requirement when entering an address or telephone number (tasks 1 and 4) does not change dramatically across sessions. Therefore, tasks 1 and 4 require fewer test sessions to obtain a
representative test time. Conversely, searching for places of interest is more varied in the input required as the participant will have to scroll through different numbers of items on the list of places before they reach the target item. Likewise, the amount through which a participant has to search through radio frequencies to find the desired radio station will vary depending on the start and end frequencies. These inconsistent tasks will require greater numbers of test sessions to obtain a representative task time.

Analysis of the results comparing the means of different sample sizes against the overall mean suggests that a minimum sample size of ten participants, each completing three task replications, would be required to obtain a reasonable representation of the performance of the population for the tasks involving programming of the navigation system (tasks 1 and 2). For task 3 (changing the radio frequency), a minimum of fourteen participants completing three task replications would be required. However, for task 4 (dialling a telephone), only seven participants completing three task replications would be required. It should be noted that an alternative to increasing the sample size would be to increase the number of test sessions completed and this is likely to be less time consuming and expensive. Although these experiments can not be used as a basis for statistically robust conclusions on the trade-off between sample size and test sessions (number of participants and number of measurements), they can be used as a basis for a reasonable practical compromise. Using engineering judgement, rather than statistical evidence, a practical recommendation would be 10 participants, each completing 5 trials.

Errors on the IVIS tasks proved to be of little value in analysis and would therefore not be a useful measure of when participants have reached a constant level of performance. However, the error data did show that participants who trained wearing occlusion goggles tended to make the same number of mistakes as participants who undertook static training. In test sessions, all participants had periodically occluded vision. If one considers that participants who completed the training tasks with occlusion goggles produced both the same number of errors and not significantly different task completion times as participants who trained without goggles, it suggests that occlusion goggles are not strictly necessary during the IVIS training period.
6 Conclusions

6.1 The IVIS tasks

- The only task considered acceptable based on existing guidelines (JAMA) was dialling a familiar telephone number in a legally installed handset. However, a suitable limit with this occlusion protocol remains to be agreed.
- Entering a destination was the task identified as requiring the most visual attention.
- Destination entry and dialling a phone number are mainly visual tasks.

6.2 The method

- Five training sessions would be required for complex tasks and fewer would be needed for less complex or familiar tasks. Task complexity is important when choosing an appropriate number of training sessions.
- IVIS task training can be conducted with or without the occlusion goggles.
- Inconsistent tasks will require a greater number of test sessions to obtain a representative task time.
- An appropriate sample size is dependent on the variability of the task. For the tasks investigated an appropriate sample size would be 14. Alternatively a smaller sample size could be used and the number of test sessions could be increased. Five test sessions with 10 participants appears reasonable (although not directly supported by statistical rigour).
- It is useful to remove system response delay from Total Task Times or a proportion of system response delay from Total Shutter Open Times, but the addition of a correction factor for short “check glances” is less useful.
- Errors were not a useful measure of visual loading. The pre-ISO protocol states that training should be conducted by participants until error free performance is reached. However, these results suggest that this would not be appropriate for many IVIS tasks in which few errors are ever made.
- TSOT appears to be a good indicator of the visual loading imposed by an IVIS task.
- The TSOT/TTTstatic ratio appears to be a good indicator of the amount of visual loading imposed by a task in relation to other sensory channels.
7 Recommendations for the protocol

<table>
<thead>
<tr>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five training sessions should be completed</td>
</tr>
<tr>
<td>Five test sessions should be completed</td>
</tr>
<tr>
<td>A sample size of 10 should be used</td>
</tr>
<tr>
<td>TSOT and the ratio TSOT/TTT\text{static} should be used as a measure of visual loading</td>
</tr>
<tr>
<td>Errors should not be used as a measure of performance</td>
</tr>
</tbody>
</table>
8 Implications of recommendations for the ISO standard

Following this study, a three day meeting of the ISO standards group TC22 SC13 WG8 took place in June 2004 in Suresnes, France; the first day of which was dedicated to discussing the Occlusion protocol. There were five presentations of ongoing work (from France, Canada, Germany and two from USA). The experimental work and conclusions from this UK research were also presented and it was appreciated that the work had tackled some of the unresolved issues with occlusion.

There then followed technical discussions to agree some important parameters:

**Removal of System Response Delay:** This was agreed using a method proposed by TRL.

**Number of training trials:** It was agreed that the goal is for participants to be familiar with the task, but not an expert. It was also agreed that a fixed number of training trials is simpler than training to some level of performance. Five training trials were agreed (as recommended in this work).

**Number of trials:** The number of trials was agreed as five (as recommended in this work).

**Number of subjects:** This was agreed as 10 (as recommended in this work).

It was accepted that the major issues had been resolved and that statistical data representation issues could be re-visited in later drafts. The group therefore agreed to finalise a “Committee Draft” to be submitted for external comment and voting. The draft refers directly to the experimental work described in this report.

The reference for this committee draft is ISO TC 22/SC 13N 763 Road vehicles – Ergonomics aspects of transport information and control systems – Occlusion method to assess visual distraction due to the use of in-vehicle systems.

Two statistical issues remained relatively open:

- The first was the best method to calculate the interruptibility parameter R which is a ratio of two other measurements. Essentially the question was whether it was better to use a ratio of the means (or medians) of the two measurements or to calculate individual ratios and then take a mean (or median).

- A second issue was how to combine all the participant data into single measures of TSOT (representing quantity of information) and R (representing interruptibility). Candidates discussed included mean, median, and 85th percentile.

As a result of this meeting, further analysis of the results obtained in this experimental work was conducted in an attempt to resolve these issues. This analysis is presented in Appendix B which is contributing to the ongoing development of the standard within WG8.
9 Summary and Recommendation

The work presented in this report has contributed to the definition of a protocol for occlusion. Using this protocol an IVIS task can be assessed for both the time it takes to complete (TSOT) and the ease with which the task can be interrupted and resumed (R - the smaller the R value, the more interruptible and resumable the interaction is). This is shown in Figure 15.

![Figure 15 IVIS task scoring](image)

As shown in Figure 15, if an IVIS task has a long task duration and is hard to resume then it would be considered unacceptable for completion while driving.

However, if an IVIS task has a long task duration and is easy to resume then it can be considered more acceptable for completion while driving.

Similarly, if an IVIS task is hard to resume and has a short task duration then it can be considered more acceptable for completion while driving.

Suitable limits for TSOT and R will not be defined by this protocol and are considered to be issues of policy. The recommended next step is to gather results for tasks currently considered acceptable and unacceptable in order to assist policy makers in identifying the acceptable and unacceptable levels of TSOT (Total Shutter Open Times) and R (Resumability Index).

With the occlusion protocol now defined, it would be relatively straightforward to measure a number of specific tasks for the purpose of setting reference values for acceptable interface designs.

Once these limiting values are defined, the occlusion protocol can be used as an objective, quick and low cost method of assessing an IVIS task.
Acknowledgements

The work described in this report was carried out in the TSU Group of TRL Limited. The authors are grateful to Alan Stevens who carried out the quality review and auditing of this report.

Glossary

Chunkable task  A task that can be broken down into small parts. If the task can be carried out in short bursts of visual attention then it is regarded as "chunkable".

Glance  Can be defined (ref.: ISO 15007, SAE J2396) as the time from the moment at which the direction of gaze moves to a target to the moment it moves away from that target. This includes the transition time to or from the target (but not both) and the dwell time on the target.

IVIS  In Vehicle Information system

Occlusion interval  The time for which the driver interface is not visible when using an occlusion procedure. During this time it is assumed that the visual attention is focussed on the driving scene (i.e. when the goggles are closed or display is off), and is also referred to in the literature as the ‘goggle close’ or ‘display off’ interval. The driver interface includes the visual display and any relevant controls.

pseudo-random  One of a sequence of numbers generated by some algorithm so as to have an even distribution over some range of values and minimal correlation between successive values.

System  Includes all components with which the manufacturer intends the driver to interact whether stand-alone or integrated into another system.

System response delay (SRD)  Interval during which the driver must wait for the driver interface to respond in order to complete a task. For example waiting for an off-board computer to be queried or waiting for a voice message to be generated.

Task  The process of achieving a specific goal using a definite and prescribed method. Example: Obtaining guidance by entering a street address using the scrolling list method until route guidance is initiated. A goal is defined as a system state sought by a driver.

TSOT – (Total Shutter Open Time)  Total time that vision is not occluded when using an occlusion procedure. TSOT is the sum of vision intervals required for the task of interest.

TTT – (Total task time)  Total time required to complete a task.

TTTstatic  Total time required to complete a task not undertaken while driving or using occlusion.

Vision interval  The periods during the occlusion technique where a participant can complete the secondary task (i.e. IVIS task). It is also referred to in the literature as the ‘goggle open interval’ ‘shutter open time’, ‘display on’ or ‘inspection interval’.
References


Society of Automotive Engineers (2000), SAE Recommended Practice Navigation and Route Guidance Function Accessibility While Driving (SAE 2364)


ISO/CD 16673 (N763R) Road vehicles – Ergonomic aspects of transport information and control systems – Occlusion Method to Assess Visual Distraction Due to the Use of In-Vehicle Systems (Current on 16th July 2004)
Appendix A  Participant instructions and tasks

Occlusion Participant Task Instructions

For the start of all tasks set the in-car display screen to the main menu page by turning on the ignition and pressing the ‘menu’ button beside the display

Task 1

Enter an address into a navigation system

1. Use the left hand rotary dial to highlight ‘GPS Navigation’ and push in to select
2. You will be asked to enter a city and street name
3. Use the rotary dial to highlight the ‘city/town’ option and push to select
4. Using the rotary dial, highlight each required letter and push to select until the city name has been spelt out (Note: the system is intuitive and will try and complete city names for you that are stored in the GPS database)
5. Note the space and delete buttons that can be highlighted and selected in the same way as for entering letters where required
6. Once the city name is complete use the rotary dial to highlight ‘enter’ and push to select
7. Once the city name has been selected, use the rotary dial to select ‘street’ and push to select
8. Enter the street name in the same way as for the city name
9. Highlight and select ‘enter’ to confirm the address
10. Task is complete when the destination has been selected and the rotary dial in pushed in to select that address

Task 2

Find information on current location

1. Use the left hand rotary dial to highlight ‘GPS Navigation’ and push in to select
2. Use the rotary dial to highlight ‘information’ and push in to select
3. Use the rotary dial to highlight ‘information on current location’ and push in to select
4. Use the rotary dial to highlight ‘the required information type (e.g. petrol station)’ and push in to select
5. Use the rotary dial to highlight the desired location from the list of options and push in to select
6. The task is complete when the desired information type and location has been selected

Task 3

Tune a radio frequency and save it as a preset station

1. Turn on the audio system by pressing in the right hand rotary dial
2. Select the correct radio frequency by pressing the ‘AM/FM’ rocker button
3. Use the ‘< >’ rocker buttons to tune the radio to the desired frequency
4. Press the desired preset number and hold until confirmation has been provided by the system that that station has been stored in the memory

Task 4

Enter a number into a mobile telephone

1. Use the numbered keys on the telephone to enter the telephone number you wish to dial
2. When the number is complete and correct press the ‘red, No’ button to signal that the task is complete
Occlusion Participant Tasks

For the start of all tasks set the in-car display screen to the menu page by turning on the ignition and pressing the ‘menu’ button beside the display

**Task 1**

Enter an address into a navigation system
Training Addresses:
- Birmingham, Duchess Road
- Manchester, Sale Road
- Reading, London Road, Reading
- Fleet, Hampshire, Dukes Mead
- Hook, Hampshire, Wild Herons

Trial Address:
- London, Churchill Gardens
- Edinburgh, Princes Street
- Farnborough, Hampshire, Westglade

**Task 2**

Find information on current location
Training Information Points:
- Bracknell Sports & Leisure Centre, Bracknell, Berkshire
- Frimley Railway Station, Camberley, Surrey
- College/University of Surrey, Guildford, Surrey
- Newbury Bus Station, Newbury, Berkshire
- Heathrow Airport, Terminal 4

Trial Information Points:
- Ciao Ninety Restaurant, Ascot, Berkshire
- BP Petrol Station, Bracknell Road
- Coppid Beach Hotel, Bracknell, Berkshire

**Task 3**

Select a radio frequency and save it as a pre-set radio station
Training System Frequencies:
- 97-99 FM Radio 1 to pre-set 1
- 95.8 FM Capital Radio to pre-set 2
- 909 MW Radio 5 to pre-set 1
- 693 MW Radio 5 to pre-set 4
- 87-91 FM BBC Radio 2 pre-set 3

Trial System Frequencies
- 1215 MW to pre-set 6
- 95.2 FM BBC Bristol to pre-set 3
- 1053 MW to pre-set 5

**Task 4**

Enter a number into a mobile telephone
- The participant should enter a number they know
Appendix B  Further analysis on different methods to Calculate R

R is a measure of the ‘chunkability’, “interruptibility” or “resumability” of an in-vehicle task and is calculated as $R = \frac{TSOT}{TTT}$.

This analysis explores the different methods of calculating and representing $R$ for a given task, using previous the data collected during the trial presented in this report for the four tasks:

- Task 1: Destination entry
- Task 2: Place of Interest (PoI) search and entry
- Task 3: Change radio frequency
- Task 4: Dial telephone number

**TTTstatic (Total Task Time static)**

For the purpose of this analysis, the last three static training sessions in which no learning effect was observed is used to provide the TTTstatic data. These three sessions, conducted by 10 participants, are referred to as X, Y and Z respectively. Tasks 1 and 2 involved a System Delay and this has been subtracted (using the ISO proposed method). The three TTTstatic times for X, Y, and Z are shown in table 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Mean</td>
<td>49.4</td>
<td>24.6</td>
<td>34.8</td>
<td>7.52</td>
</tr>
<tr>
<td>Median</td>
<td>46.7</td>
<td>19.2</td>
<td>32.7</td>
<td>7.25</td>
</tr>
<tr>
<td>Min</td>
<td>30.2</td>
<td>1.10</td>
<td>11.0</td>
<td>4.40</td>
</tr>
<tr>
<td>Max</td>
<td>110.8</td>
<td>110.8</td>
<td>84.5</td>
<td>16.9</td>
</tr>
<tr>
<td>SD</td>
<td>19.5</td>
<td>20.6</td>
<td>14.0</td>
<td>2.43</td>
</tr>
<tr>
<td>85th %ile</td>
<td>59</td>
<td>36</td>
<td>46</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Table 1 TTT times for each IVIS task

**TSOT (Total Shutter Open Time)**

After training, the 10 participants completed 3 test trials of each task whilst wearing occlusion goggles. The test trials, 1, 2, and 3 gave three TSOTs and these are referred to as A, B, and C respectively.

TSOT times using task times for test trials A, B, and C (times for tasks 1 & 2 have had system delay subtracted) are shown in table 2.

<table>
<thead>
<tr>
<th>Task</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Mean</td>
<td>51.1</td>
<td>17.2</td>
<td>18.1</td>
<td>6.69</td>
</tr>
<tr>
<td>Median</td>
<td>47.4</td>
<td>15.5</td>
<td>16.0</td>
<td>5.00</td>
</tr>
<tr>
<td>Min</td>
<td>31.4</td>
<td>9.20</td>
<td>8.40</td>
<td>2.40</td>
</tr>
<tr>
<td>Max</td>
<td>100.1</td>
<td>58.8</td>
<td>48.9</td>
<td>50.4</td>
</tr>
<tr>
<td>SD</td>
<td>16.5</td>
<td>9.25</td>
<td>8.76</td>
<td>8.44</td>
</tr>
<tr>
<td>85th %ile</td>
<td>66</td>
<td>21.5</td>
<td>27</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Table 2 TSOT times for each IVIS task
R

From the data collected, calculation of R can be performed in a number of ways and three are investigated here.

(a) R using paired training (TTT) and test (TSOT) times

R can be calculated using nominal ‘pairs’ of trials, i.e. three values of R could be calculated for each subject using the three TSOT and TTT values recorded for each task e.g.:

\[ \text{Subject 3, Task 2: } R_1 = \frac{A}{X}, \quad R_2 = \frac{B}{Y}, \quad R_3 = \frac{C}{Z} \]

This gives a value of R for each pair of TSOT and TTT values in each task. The resultant R values generated for each task when R is calculated in this way are shown in table 3.

<table>
<thead>
<tr>
<th>Task</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Mean</td>
<td>1.14</td>
<td>1.87</td>
<td>0.61</td>
<td>0.95</td>
</tr>
<tr>
<td>Median</td>
<td>1.01</td>
<td>0.71</td>
<td>0.47</td>
<td>0.74</td>
</tr>
<tr>
<td>Min</td>
<td>0.49</td>
<td>0.12</td>
<td>0.12</td>
<td>0.24</td>
</tr>
<tr>
<td>Max</td>
<td>3.03</td>
<td>21.3</td>
<td>1.91</td>
<td>8.26</td>
</tr>
<tr>
<td>SD</td>
<td>0.55</td>
<td>4.03</td>
<td>0.41</td>
<td>1.39</td>
</tr>
<tr>
<td>85th %ile</td>
<td>1.50</td>
<td>1.75</td>
<td>0.85</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Table 3 R values calculated using paired training and test times

The mean R values show that the PoI search (Task 2) is the most visually demanding task. However, closer inspection of the data shows that the 85th percentile value is less than the mean value, suggesting that the mean is being skewed by one or more particularly high values of R. The maximum R value for task 2 is 21.3 relative to a mean of 1.87. It is counterintuitive that the TSOT time should be more than twenty times greater than the TTT time and suggests an abnormal result.

(b) R using the respective means of (TTT) and (TSOT) times per subject

Alternatively, R can be calculated on a per subject basis by taking the mean TSOT time across the three test repetitions for each task and the mean TTT time across the three training repetitions for each task e.g.:

\[ \text{Subject 3, Task 2: } R = \frac{\text{Mean}(A, B, C)}{\text{Mean}(X, Y, Z)} \]

Calculating R in this way only gives one value of R for each subject in each task. The values calculated for each subject using the mean of the three values of TSOT and TTT respectively are shown in table 4.

<table>
<thead>
<tr>
<th>Task</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Mean</td>
<td>1.10</td>
<td>0.88</td>
<td>0.53</td>
<td>0.90</td>
</tr>
<tr>
<td>Median</td>
<td>0.99</td>
<td>0.74</td>
<td>0.55</td>
<td>0.71</td>
</tr>
<tr>
<td>Min</td>
<td>0.76</td>
<td>0.20</td>
<td>0.32</td>
<td>0.44</td>
</tr>
<tr>
<td>Max</td>
<td>2.47</td>
<td>2.18</td>
<td>0.66</td>
<td>2.87</td>
</tr>
<tr>
<td>SD</td>
<td>0.50</td>
<td>0.59</td>
<td>0.13</td>
<td>0.71</td>
</tr>
<tr>
<td>85th %ile</td>
<td>1.09</td>
<td>1.30</td>
<td>0.65</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 4 R values calculated using the mean training and test times per subject
Using this method for the calculation of R, the destination entry task (Task 1) now has the highest mean R value. The PoI search task (Task 2) has the highest 85th percentile R value (R_{85}). The maximum R values are now much more reasonable and in general, this formulation appears to be much more representative than if single pairs of TSOT and TTT values are used. Although the number of data points is reduced by taking the mean (N = 10 instead of N = 30), this reduces the likelihood that the R value for a task will be skewed by an atypical ‘pair’ of TSOT and TTT task times as observed in the 3(a).

(c) R using the respective medians of (TTT) and (TSOT) times per subject

A further alternative using a similar technique is to calculate R for each subject using the median TSOT time of the test repetitions for each task and the median TTT time of the training repetitions for each task e.g.:

\[
R = \frac{\text{Median}(A, B, C)}{\text{Median}(X, Y, Z)}
\]

As in (b), this gives one value of R for each subject in each task. The values calculated for each subject using the median of the three values of TSOT and TTT respectively are shown in table 5.

<table>
<thead>
<tr>
<th>Task</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Mean</td>
<td>1.04</td>
<td>0.84</td>
<td>0.52</td>
<td>0.71</td>
</tr>
<tr>
<td>Median</td>
<td>0.98</td>
<td>0.75</td>
<td>0.48</td>
<td>0.72</td>
</tr>
<tr>
<td>Min</td>
<td>0.59</td>
<td>0.26</td>
<td>0.30</td>
<td>0.49</td>
</tr>
<tr>
<td>Max</td>
<td>2.31</td>
<td>1.47</td>
<td>0.74</td>
<td>0.95</td>
</tr>
<tr>
<td>SD</td>
<td>0.49</td>
<td>0.41</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>85th %ile</td>
<td>1.09</td>
<td>1.30</td>
<td>0.72</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 5 R values calculated using the median training and test times per subject

As before in method (b), the destination entry task (Task 1) has the highest mean R value. The PoI search task (Task 2) has the highest 85th percentile R value (R_{85}). By taking median values, this technique further reduces the possibility of obtaining a biased result caused by anomalous TSOTs or TTTs. The standard deviation of the R values is now less than it was in method (b) for three out of the four tasks, indicating more consistent data.

It is therefore recommended that all measures of the value of R are calculated on a within-subject basis by taking the ratio of the respective median TSOT and TTT times across trial repetitions completed by one subject on each task under evaluation. Mean (R_{\text{mean}}), median (R_{\text{median}}), or 85th percentile (\text{R}_{85}) R values can then be taken across subjects for each task. This technique appears to give a better estimate of the R values for each task.

Finally, whether a mean, median, min, max or 85th percentile value of R is chosen as the single R value to represent the task is a value judgement that depends on the purpose of the measurement. For the purposes of identifying interface designs that exhibit poor interruptibility, the R_{85} seems most appropriate. This represents a “reasonable worst case” from the data without taking extreme values.